

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-123465

(43)Date of publication of application : 17.05.1996

(51)Int.Cl.

G10L 3/00

G10L 3/00

G10L 3/00

(21)Application number : 06-264097

(71)Applicant : NIPPON TELEGR &amp; TELEPH CORP &lt;NTT&gt;

(22)Date of filing : 27.10.1994

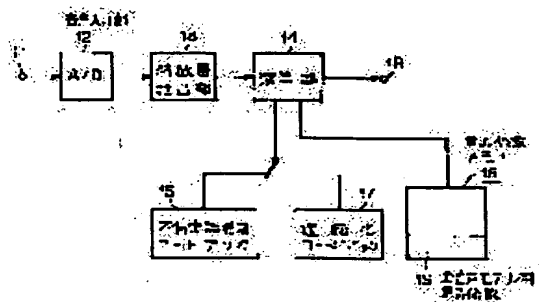
(72)Inventor : MATSUOKA TATSUO  
FURUI SADAHIRO

## (54) ADAPTING METHOD FOR ACOUSTIC MODEL

## (57)Abstract:

**PURPOSE:** To increase a recognition rate with a small learning speech and a small calculation quantity.

**CONSTITUTION:** A semicontinuous distribution HMM model is generated by using a learning speech for an unspecified speaker, its base distribution is stored in a code book 15, and weight coefficients for respective base distributions are stored in a weight coefficient memory 16; and weight coefficients as to all phonemes independent of respective phonemes are stored as weight coefficients 19 for all phoneme models. The learning speech of a recognition speech is inputted, the weight coefficients 19 for all the phoneme models are used, and only the respective base distributions in the code book 15 are adapted and stored in a code book 17. At the time of recognition, an input sound is recognized by using the weight coefficients in the code book 17 and the weight coefficients in the weight coefficient memory 16.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the  
examiner's decision of rejection or application  
converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of  
rejection][Date of requesting appeal against examiner's decision  
of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平8-123465

(43)公開日 平成8年(1996)5月17日

(51)Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 3/00	5 2 1 D			
	5 3 1 K			
	G			
	5 3 5			

審査請求 未請求 請求項の数3 O L (全 9 頁)

(21)出願番号	特願平6-264097	(71)出願人	000004226 日本電信電話株式会社 東京都新宿区西新宿三丁目19番2号
(22)出願日	平成6年(1994)10月27日	(72)発明者	松岡 達雄 東京都千代田区内幸町1丁目1番6号 日 本電信電話株式会社内
		(72)発明者	古井 貞照 東京都千代田区内幸町1丁目1番6号 日 本電信電話株式会社内
		(74)代理人	弁理士 草野 卓

(54)【発明の名称】 音響モデルの適応化法

(57)【要約】

【目的】 少ない学習音声、少ない計算量で認識率を高くする。

【構成】 不特定話者用学習音声を用いて、半連続分布HMMモデルを作り、その基底分布をコードブック15に記憶しておき、各音素についての、各基底分布に対する重み係数を重み係数メモリ16に記憶しておき、また各音素とは独立の全音素についての重み係数を全音素モデル用重み係数19として記憶しておく、認識音声の学習音声を入力して、全音素モデル用重み係数19を用いて、コードブック15の各基底分布のみを適応化してコードブック17を格納する。認識時には入力音響をコードブック17と重み係数メモリ16の重み係数とを用いて認識する。

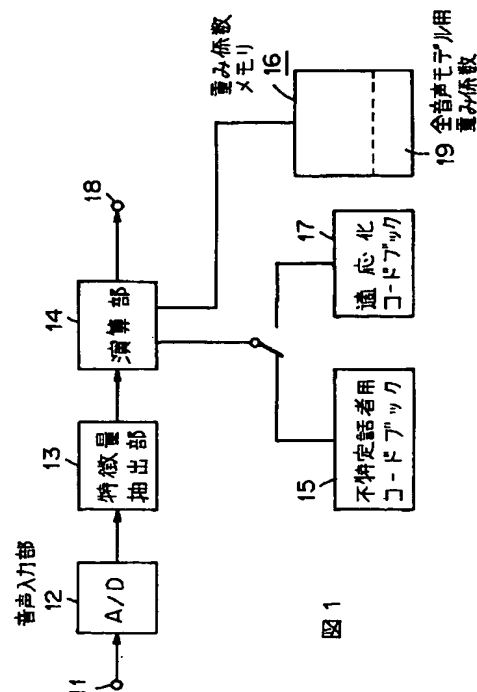


図 1

## 【特許請求の範囲】

【請求項1】 学習用音声を用いて、その音声の音響的特徴量を抽出し、その特徴量を統計的にモデル化して、認識カテゴリに対応した音響モデルを、認識時に、上記学習用音声と性質を異にする音声を用いて適応化する方法において、

上記音響モデルは、パラメータ空間を複数の基底分布で表現したコードブックと、そのコードブック中の各基底分布に対する重み係数とにより構成し、

各認識対象カテゴリと独立に学習した全カテゴリ音響モデルを用いて、上記パラメータ空間を表現する基底分布を、上記性質を異にする音声により再推定して適応化することを特徴とする音響モデルの適応化法。

【請求項2】 上記再推定された各基底分布の、その推定前の基底分布に対する変化をそれぞれ適応化ベクトルとし、

上記基底分布を音声パワーに従ってクラスタリングし、上記適応化ベクトルを上記各クラスに属する基底分布について平均化し、

その平均化適応化ベクトルを用いて、そのクラスターの各基底分布を適応化することを特徴とする請求項1記載の音響モデルの適応化法。

【請求項3】 上記再推定された各基底分布の、その推定前の基底分布に対する変化をそれぞれ適応化ベクトルとし、

上記基底分布を音声パワーに従ってクラスタリングし、上記適応化ベクトルを上記各クラスに属する基底分布について平均化し、

その各クラスごとの平均化適応化ベクトルと、そのクラスターの各基底分布ごとの適応化ベクトルとを荷重平均し、

その荷重平均適応化ベクトルを用いて、そのクラスターの基底分布を適応化することを特徴とする請求項1記載の音響モデルの適応化法。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】この発明は、音声認識における標準パターンとなるべきものとして用いられ、あらかじめ別の環境で収録された音声（学習用音声）を用いて学習した音響モデルを、特定の音声収音系回線特性や、特定の話者など学習用音声と性質を異にする音声に適応化する方法に関する。

## 【0002】

【従来の技術】音声の音響的特徴を確率的、統計的にモデル化する手法である隠れマルコフモデル（Hidden Markov Model：HMM）を用いた音声認識システムでは、一認識対象カテゴリ、つまり音素、音節、単語などの語彙（あるいは認識対象単位）ごとに、1つ、あるいは複数のHMMを設定し、学習用音声を用いて学習する、つまりHMMを作る。認識時には、

音声認識システムの入力音声それぞれのモデルから観測される確率を計算し、尤度（尤もらしさ）の最も高い順に認識結果候補としている。HMMは、統計的なモデルであるから学習用音声中に現われた頻度に従って、ある音響的特徴量とあるカテゴリとを関連づける強さを内部に確率分布として表現する。つまり図4Aに示すように、すべての認識対象カテゴリ（例えば音素）ごとに、初期状態（音素の始端付近）a、第2状態b、第3状態c、最終状態（音素の終端）dの4つの状態を順次遷移し、各状態はその音素のその状態における音響特徴量の統計的な分布を表現し、状態から状態への遷移確率が与えられた音響モデル $M_1 \sim M_n$ が予め求められ、入力音声がある音響モデルより出力する確率を計算して入力音声に対するその音響モデルの尤度を求める。

【0003】HMMは確率分布の表現方法から、離散分布モデル、連続分布モデル、半連続分布モデルの3つに大きく分類される。離散確率分布モデルでは、音声の音響的特徴量はコード化された離散的な値で表現される。例えば、図5Aに示すように音声の音響的特徴量は代表的なN個の特徴ベクトル $A_1 \sim A_n$ の何れかで表わされ、これら特徴ベクトル $A_1 \sim A_n$ にはそれぞれコード（例えば番号） $C_1 \sim C_n$ が与えられている。また各音素を示す音響モデル $M_1 \sim M_n$ のそれぞれごとに、図5Bに示すように、コード $C_1 \sim C_n$ のそれぞれに対し、出力確率 $P_1 \sim P_n$ が1対1で対応づけられている。入力音声はフレームごとにその特徴ベクトルが代表特徴ベクトル $A_1 \sim A_n$ の何れに最も近いかが求められ、その代表特徴ベクトルを示すコード列に入力音声に変換され、そのコード列は各音響モデル $M_1 \sim M_n$ のそれぞれについて、その出力確率が演算される。これら演算された出力確率中の最も高い（尤度が大きい）音響モデルと対応する音素が認識結果として出力される。

【0004】連続確率分布モデルでは、音響的特徴量ベクトルは連続量のまま扱う。例えば図4A中の音響モデル $M_n$ はその初期状態aはその音響的特徴量が分布 $D_1$ で表わされ、状態b～dではそれぞれその音響的特徴量が分布 $D_2 \sim D_4$ として表わされる。連続確率分布モデルには、単一分布モデルと混合分布モデルとがあり、図4Aのモデル $M_n$ は混合分布モデルの場合で例えば図4Bに示すように、1つの混合分布 $D$ が複数の分布 $V_1 \sim V_r$ の重み付き加算の形で表現される。これら分布 $V_1 \sim V_r$ は音声の音響的特徴量の分布をガウス分布で近似し、平均値 $\mu_1 \sim \mu_r$ と共分散行列 $\sigma_1 \sim \sigma_r$ とに止りそれぞれ表現される。音響モデル $M_1 \sim M_n$ それぞれ図5Cに示すように各状態ごとにその複数の分布と、図に示していないが重み係数とが与えられて表現される。

【0005】入力音声は各音響モデルごとに、その表現された各状態の分布により出力確率を演算し、音響モデルごとの出力確率、つまり尤度を求めその最大の音響モデルの音素を認識結果とする。混合分布モデルは精密な

分布の推定が可能であるが、推定すべきパラメータ数が多いため、それだけ多くの学習用音声が必要とする。半連続確率分布モデルは離散分布モデルと連続分布モデルの混合分布形のものの特徴を合せたモデルである。つまり混合ガウス分布の連続分布モデルにおいて、混合分布数を十分大きく、例えば256に設定し、かつ各音響モデルに対して同一の分布 $V_1 \sim V_N$ とし、各音響モデル間の区別は重み係数によって行う。例えば図6に示すように状態aについて、各音響モデル $M_1 \sim M_N$ について、分布 $V_1 \sim V_N$ のそれぞれに対する重み $W$ がそれぞれ与えられている。同様に状態b、c、dについても、各音響モデル $M_1 \sim M_N$ のそれぞれに対し、分布 $V_1 \sim V_N$ のそれぞれの重み $W$ が与えられている。つまり基底分布 $V_1 \sim V_N$ は全音響モデル、全状態にわたって共有され、各音響モデルの各状態ごとに重み係数 $W_i$ の値が各音素固有の値として決められている。入力音声は各音響モデルごとに、出力確率を演算し、その最大の音響モデルの音素を認識結果とする。半連続モデルは離散モデルにおける特徴ベクトル $A_1 \sim A_N$ の代りに基底分布 $V_1 \sim V_N$ が用いられたもので、図5Aに示した単一のコードブックによりパラメータ空間が表現されているという離散モデルの特徴と、混合ガウス分布により各音素モデルが詳細に表現されるという混合分布形連続モデルの特徴とを合せもっている。

【0006】HMMのような統計的なモデルを用いた音声認識では、モデルパラメータを推定するための学習用音声と実際に認識対象になる音声とが同じような条件で收音されることを前提としている。すなわち、音響的な環境、たとえば、背景雑音や、回線特性が、学習時と認識時ではほぼ同じであると仮定している。学習時と認識時との收音条件が異なる場合、実際に認識対象となる音声の音響的特徴量はモデルが表現している音響的特徴量と異なるので認識精度が悪くなるという問題がある。

【0007】学習時と認識時との音響的特徴量の変動には、スペクトル上で加算的に影響するものと、フィルタ的に影響するものがある。背景雑音などはパワーとして音声に加わるものであるからスペクトル領域でも加算的になる。一方、回線特性の違い(歪み)などはスペクトル包絡の形状が変化、通常はスペクトル包絡の傾きが変化するのでスペクトル領域においてフィルタ的に影響する。

【0008】学習時と認識時との音響的な条件が異なる場合、認識システムを認識対象となる音響的条件に適応化することで認識性能を改善しようとする試みがされてきた。以下に、これまでに提案されている2つの方法について説明する。第一は、ケプストラム平均値正規化法と呼ばれる方法である。音声の音響的特徴量としては対数スペクトルの逆フーリエ変換で定義されるケプストラムが用いられることが多い。ケプストラム領域においては、スペクトラム領域におけるフィルタが加減算により

実現されるので回線特性の変動による歪みはケプストラムの加減算により補正できる。この原理による簡単で効果的な回線特性補正方法がケプストラム平均値正規化法である。音声の音響的特徴量としてケプストラム係数を用いる場合には、そのケプストラムの時系列から当該音声区間にわたる平均値を引くことで、時不変な周波数スペクトルの傾向を平坦化することができる。しかしながら、ケプストラム平均値正規化法では、長時間平均により回線における時不変のスペクトル包絡を差し引いて平坦化するというのが原理であるため、ある程度長い音声区間にわたって平均をとらないと効果が期待できない。また、単純にある区間のケプストラムの時系列の平均値を差し引くだけであるため、音声エネルギーの大小、あるいはSN比の違いによる影響で推定誤りが起こるなどその改善効果には限界があることが問題であった。

【0009】第二はコードブックの変換によるモデル適応化法である。この方法は、話者適応化のために提案されたが、コードブックを用いるモデルをベースとしていれば、一般に学習音声と認識対象音声の収録環境の不一致に対する適応化手法として適用可能と考えられる。この方法により、離散確率分布モデル、あるいは半連続分布モデルの場合には、コードブックを学習用音声で求めたものから認識対象となる音声で求めたものへ変換することでモデルの適応化が可能である。この方法について、学習用音声の収録回線である回線Aの音声で学習したモデルを、認識対象音声の収録回線である回線Bの音声に適応化する場合を例として説明する。回線Aの音声と回線Bの音声とがあるとき、回線Aの音声を用いてコードブックAを、回線Bの音声を用いてコードブックBをそれぞれ設計する。そして、回線Aの音声をコードブックAを用いてベクトル量子化し、その結果のコードブックAのコードの系列を用いてHMMを学習する(HMMを作成する)。次に、発声内容が同じ回線Bの音声を、それぞれコードブックA、コードブックBを用いてそれぞれベクトル量子化し、コードブックAとコードブックBの各コードの対応関係をDPマッチングにより求める。回線Bの音声を認識対象とするときには、コードブックBでベクトル量子化を行ない、その結果をコードブックAとコードブックBの対応関係からコードブックAのコード系列に変換し、コードブックAを用いて学習したHMMを用いて回線Bの音声を認識することが可能になる。しかしながら、この方法は回線Bの音声、すなわち、認識対象となる音声の収録された回線の音声をコードブックを設計できるほどの量を持っていることが必要であり、かつ回線Aと全く同じ発声内容の音声が必要でなければならぬということが問題である。したがって、より少ない量の適応化音声で、かつ発声内容に関する制約の緩い適応化法が必要であった。

【0010】

【発明が解決しようとする課題】この発明の目的は、学

習用音声と認識対象となる音声とが性質の異なる場合にも高い認識精度を得るための適応化を、少ない学習音声を用いて行うことができ、また少ない計算量で行なうことができる音響モデルの適応化方法を提供することにある。

【0011】

【課題を解決するための手段】この発明によれば音響モデルを、パラメータ空間を複数の基底分布で表現されたコードブックと、そのコードブック中の各基底分布に対する重み係数とにより構成し、各認識対象カテゴリと独立に学習した全カテゴリ音響モデルを用いて、パラメータ空間を表現する基底分布を、性質を異にする音声、つまり認識時の音声と同一性質の音声により再推定して適応化する。

【0012】請求項2の発明では、再推定された各基底分布の、その推定前の基底分布に対する変化をそれぞれ適応化ベクトルとし、各基底分布を音声パワーに従ってクラスタリングし、上記適応化ベクトルを上記各クラスタに属する基底分布について平均化し、その平均化適応化ベクトルを用いて、そのクラスタの各基底分布を適応化する。

【0013】請求項3の発明では各クラスタごとの平均化適応化ベクトルと、そのクラスタの各基底分布ごとの適応化ベクトルとを荷重平均し、その荷重平均適応化ベクトルを用いてそのクラスタの基底分布を適応化する。

【0014】

【作用】前記構成により、(1)全音素HMMを用いることで適応化音声の発声内容によらずに任意の発声により適応化が可能であり、(2)音声のパワーの大小を考慮しているため、より正確な回線特性の適応化が可能であるという利点がある。すなわち、音声のパワーが大きくSN比が高い場合は付加的な雑音の影響が小さく、パワーが小さい場合はその逆であることを利用し、パワーの大きいクラスタに属する基底分布ロードワードに対してはそのクラスタに属する基底分布(コードワード)の修正量の平均値を、パワーの小さいクラスタに属する基底分布(コードワード)に対してはその基底分布自身の修正量を重視するようにコードブックを適応化することが可能である。

【0015】

【実施例】以下、この発明の一実施例として、防音室など音響条件の比較的良好な環境で収録した音声で学習した音響モデルを、学習音声とは特性の異なる電話音声に適

$$F(x) = W_1 V_1(x) + W_2 V_2(x) + \dots + W_n V_n(x)$$

で求められる。 $W_1 \sim W_n$ は各HMMにより異なった値である。

【0018】各音素と対応するHMMの形を決めるパラメータ( $V_1, V_2, V_3, \dots, V_n$ のガウス分布の平均値と共分散、それぞれの分布に対する重み係数、

\* 応化する場合について図面を参照して説明する。この例では音響モデルとして半連続分布HMMを用いた場合について説明する。この発明の方法は、モデルパラメータ空間を基底分布の集合により表現し、その基底分布を各モデルが共有するようなモデル表現であれば、離散分布HMMでも連続分布HMMでも適用可能である。

【0016】図1にこの発明を適用した音声認識装置を示す。入力端子11からのアナログ音声信号は音声入力部12でデジタル音声信号に変換され、そのデジタル音声信号から音響特徴量(例えば、ケプストラム、 $\Delta$ ケプストラム、 $\Delta$ パワーなど)が音響特徴量抽出部13で抽出される。音響モデルとしてHMMを用いた場合で、HMMのパラメータ(音響特徴量ベクトルの平均値、共分散、遷移確率)や、各分布の重み係数は演算部14で計算される。半連続分布HMMではパラメータ空間を複数の基底分布で表現したコードブックと、そのコードブック中の各基底分布に対する重み係数とにより構成されるが、前記比較的良好な環境で収録した音声で学習したHMMの基底分布が不特定コードブック14に蓄えられ、その各HMMについての各基底分布に対する重み係数が重み係数メモリ16に記憶されている。またこの発明では不特定話者用コードブックの基底分布を電話音声で適応化した基底分布が適応化コードブック17に蓄えられる。認識結果は演算部14から出力端子18に出力される。音響特徴量抽出部13は、ハードウェアにより実現しても、あるいは、ソフトウェアにより実現してもよい。ソフトウェアにより実現する場合には、演算部14の演算能力が十分にあれば演算部14で実現しても差しつかえない。

【0017】適応化前の基底分布の集合、つまり不特定話者用コードブック15に収容されている基底分布の集合は例えば図2Aに示すように $V_1 \sim V_n$ からなる。半連続HMMは前述したようにこのコードブックの各分布に対する重み係数をもっており、入力音声に対する尤度は、各分布の確率分布関数値を重み付き加算することにより求められる。コードブック15のサイズ、すなわち基底分布の数は、音響的特徴量として例えばケプストラム係数を用いる場合、256程度を用いることが多い。入力音声の特徴ベクトルを $x$ 、各基底分布の確率密度関数値を $V_1(x), V_2(x), V_3(x), \dots, V_n(x)$ とし、それぞれの分布に対する重み係数を $W_1, W_2, W_3, \dots, W_n$ とすると、その入力音声の特徴ベクトル $x$ に対する尤度 $F(x)$ は

$$\dots (1)$$

$W_1, W_2, W_3, \dots, W_n$ は、多くの音声データを用いてフォワード・バックワードアルゴリズムにより推定される。ここで、基底分布 $V_1 \sim V_n$ は全モデル、全状態にわたって共有されており、各モデルの各状態ごとに $W_i$ の値が各音素モデル固有の値として推定される。

【0019】この発明による適応化では、各モデルの $W_i$ 、つまり重み係数メモリ16の内容はそのままにしておき、基底分布 $V_1 \sim V_N$ のみを適応化する。適応化により、認識対象となる音声のパラメータ空間をうまく表現できるように各基底分布 $V_1 \sim V_N$ の平均値や分散が変化する。平均値の変化により位置が移動し、共分散の変化により分布の大きさが変わる。この適応化により各分布 $V_1 \sim V_N$ は図2Bに示すように変化させられる。適応化用音声十分な量を得られない場合には共分散は変化させずに平均値だけを変化させてもよい。基底分布 $V_1 \sim V_N$ 自体が新たなパラメータ空間へ移動することによりモデル固有の $W_i$ が変化しなくとも、音素モデルとしては新たなパラメータ空間に適応化されたものとなる。

【0020】適応化による平均値、共分散の再推定の具体的方法について、モデルを音素単位に設定していると仮定して説明する。平均値、共分散の再推定は全音素HMMを用いて行なう。つまり、各認識音素とは独立に、認識対象となる音声すべてを用いて学習し、全ての音素の何れに対しても比較的大きな尤度となるように学習したモデル、いわゆる全音素モデルに対する重み係数 $W_1 \sim W_N$ を用いて、コードブック15のみ（平均値と共分散）を再学習する。通常、各音素モデルは、256のコードワード、つまり基底分布の内、特にその音素を表現するために重要な複数の基底分布に対して高い重み係数を持ち、その他に対してはほとんど0に近い非常に小さな重み係数の値を示す。したがって、個々の音素モデルの再推定では、大きな重み係数のなかったコードワード（基底分布）がより大きく移動し、重み係数の小さなコードワード（基底分布）はほとんど移動しないため、コードブック全体をバランスよく再推定することができないため、全コードワード（全基底分布）に対してバランスよく重み係数を持っている全音素HMMを用いて再推定を行なう。この全音素HMMは不特定話者用コードブック15及び重み係数を学習する際に予め学習しておき、その重み係数をメモリ16中に全音素モデル用重み係数19として記憶しておき、適応化コードブック17を作成する際に、この全音素モデル用重み係数を用いて、その他は通常の学習と同様にフォワード・バックワードアルゴリズムにより各音素モデル（HMM）の平均値及び共分散の推定を行って適応化コードブック17を作成する。

【0021】全音素モデルは音素に独立なので発声内容によらずに学習できるため、ある決められた適応化用学習音声を発声しなければならないというような拘束条件を必要としないことも利点である。以上の適応化学習の演算は、図1中の演算部14において行なわれる。回線Bの適応化学習用音声は、各認識カテゴリ（音素）に対応する区間をラベル付けされている必要がなく、回線Bの適応化学習用音声は、音声入力部12、音響特徴量抽

出部13において、アナログ音声信号からデジタル音声信号に変換され、音響特徴量ベクトルにされる。この回線Bの適応化学習音声の音声区間の音響特徴量ベクトルを観測サンプルとして、フォワード・バックワードアルゴリズムにより全音素HMMの分布の平均値、共分散や、重み係数を再推定することができる。各音素HMMは重み係数の再推定／更新をする必要がなく、コードブック15を適応化されたものに変更するだけでよい。このようにして、基底分布の重み係数はもとの不特定話者用モデル、つまりメモリ16の内容と同じで、コードブック15の平均値、共分散が回線Bの音声に最適化されたHMMを作成し、適応化コードブック17とされる。

【0022】通常の不特定話者音声認識では、不特定話者用コードブック15と重み係数メモリ16で音響特徴量が表現された不特定話者用モデルを用いる。回線Bからの音声を認識する場合は、適応化コードブック17と重み係数メモリ16とで認識対象回線Bに適応化されたHMMを用いて、回線Bの入力音声に対する各認識カテゴリのHMMの尤度を求め、最も尤度の高いモデルのカテゴリを認識結果とする、あるいは尤度の高い順に認識結果候補とする。

【0023】図3Aに、この発明方法によりマイク音声で学習した半連続HMMのコードブック15を電話音声へ適応化した場合のその電話音声に対する音素認識結果を示す。音響的特徴量はケプストラムと $\Delta$ ケプストラム各12次元である。図中、CMNは従来技術の項で述べたケプストラム平均値正規化法、meanは各基底分布の平均値だけを適応化したもの、mean and varは平均値と共分散を同時に適応化したもの、mean+varは平均値だけを適応化した後に、共分散だけを適応化したものである。この図からCMNにこの発明方法を組み合わせると55.4%まで認識率が向上した。

#### 請求項2の発明の実施例

半連続HMMのコードブック15を各基底分布の音声パワーにしたがってクラスタリングする。すなわち音声パワーの近い基底分布は同じクラスタに属する。前記請求項1の発明の実施例において求めた各基底分布（コードブック15）に対応する適応化基底分布（コードブック17）の変化を適応化ベクトルとする時、基底分布の属するクラスタごとにその適応化ベクトルを平均化して、そのクラスタの代表適応化ベクトル（平均化適応化ベクトル）とし、そのクラスタに属する基底分布すべてをそのクラスタの代表適応化ベクトルにより適応化する。例えば音声パワークラスタリングにより、例えば図2A中の基底分布 $V_1, V_2, V_3$ が同じクラスタに属したとすると、基底分布 $V_1, V_2, V_3$ の適応化コードブック17中の各対応する基底分布への変化ベクトル（適応化ベクトル） $E_1, E_2, E_3$ （この場合は平均値の変化を示すベクトル）を平均化し、その平均化適応化ベ

トルE。を用いて、そのクラスタに属する基底分布  $V_1, V_2, V_3$  を適応化する。この場合は一種の平滑化の効果により適応化用音声が少ない場合にも頑健な適応化が期待できる。

【0024】図3Bにこの請求項2の発明の方法でコードブックを適応化した場合の認識結果を示す。音響的特徴量としてケプストラム、 $\Delta$ ケプストラムに加え、正規化対数パワーとその一次微分( $\Delta$ パワー)を用いた。クラスタリングは正規化対数パワーにより行なった。クラスタ数は実験的に最適値を求め、5とした。特徴量が増えたことにより先の実験より全体的に認識率が向上しているが、パワーでクラスタリングした場合はCMNやmean(全音素HMMでコードブックの平均値を適応化した場合)より高い認識率を示している。

#### 請求項3の発明の実施例

前記実施例における各基底分布に対応する適応化ベクトルと、その基底分布の属するクラスタの代表適応化ベクトル(平均化適応化ベクトル)との重み付き線形和を新たに適応化ベクトルとしてコードブックを適応化する。音声パワーが大きいところではおもにフィルタ的な歪みの影響が精度の劣化原因として考えられ、音声パワーが小さいところでは加算的な雑音の影響も無視できないと考えられるため、音声パワーの大小によって、基底分布自身に対応する適応化ベクトルとクラスタの代表適応化ベクトルの寄与率を操作することで、より高精度な適応化が実現できると期待できる。

\*

\*【0025】上述ではこの発明を回線音声に適応化させる場合に適用したが、いわゆる話者適応にも適用できる。また音響モデルとしてはHMMに限らない。

【0026】

【発明の効果】以上述べたように、この発明によれば、(1)任意の発声内容の適応化音声により認識対象となる音声の特性へ音響モデルを適応化することができ、

(2)音声パワーに応じた適応化を行なうことでより頑健で精度の高い適応化が可能となる、(3)各カテゴリモデルの分布係数は再推定せず、共通のコードブックだけを再推定するため適応化学習に要する学習音声は少なくよく、そのため計算時間も少ない、などの利点がある。

【図面の簡単な説明】

【図1】この発明を適用した音声認識システムの構成を示すブロック図。

【図2】この発明による音響モデルの適応化の様子を示す図。

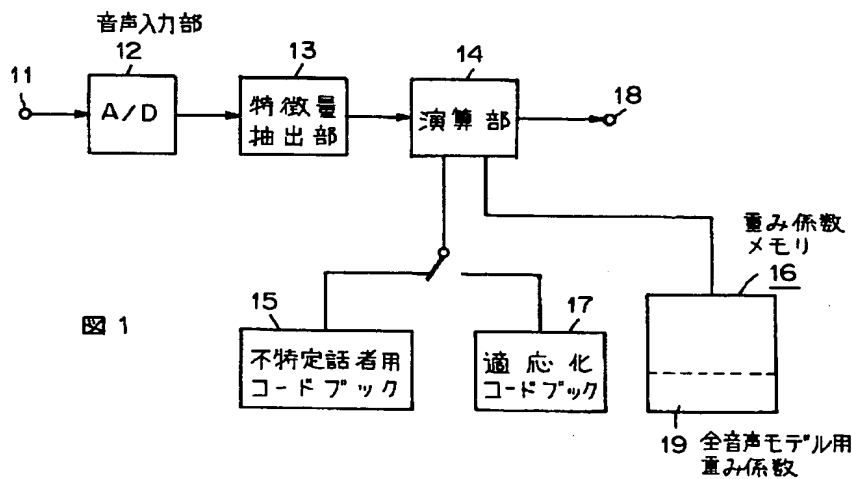
【図3】この発明の効果を示す図。

【図4】Aは音響モデルの例を示す図。Bは混合分布の例を示す図である。

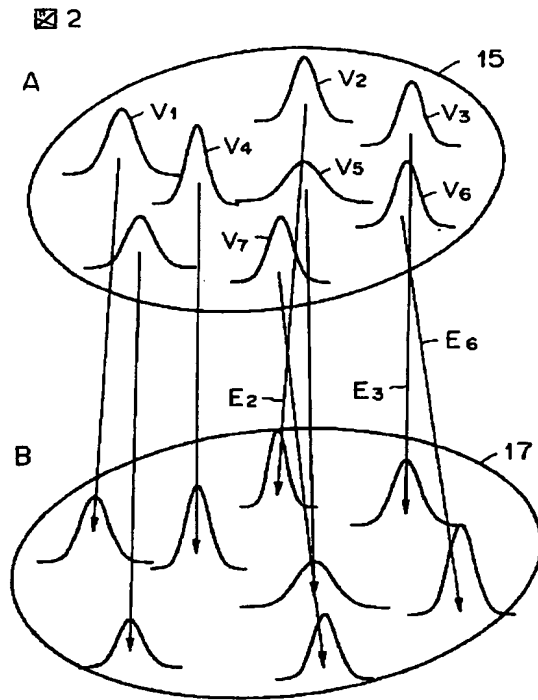
【図5】Aは離散分布モデルのコードブックの例を示す図、Bはその各音響モデルの例を示す図、Cは連続分布モデルの例を示す図である。

【図6】半連続分布モデルの例を示す図。

【図1】

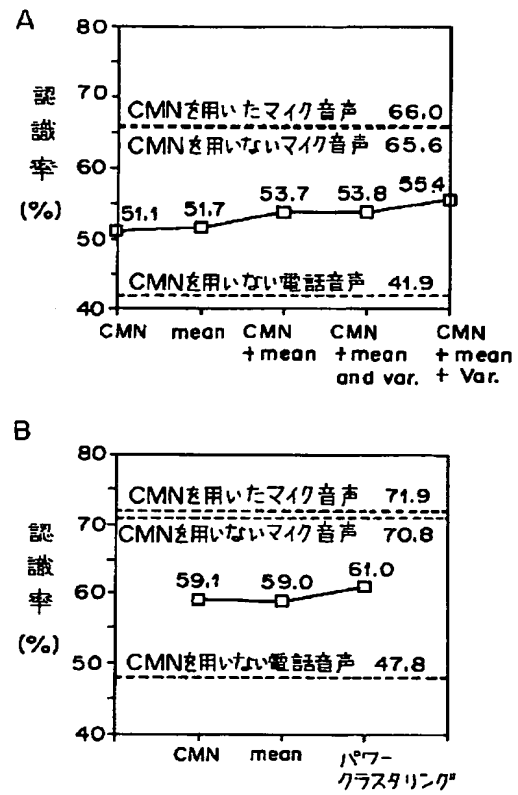


【図2】



【図3】

図3

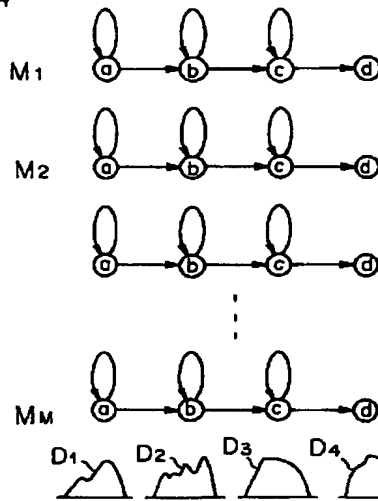




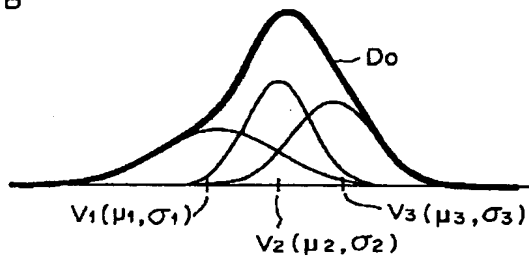
【図4】

図 4

A



B



【図5】

図 5

A

特徴ベクトル	コード
$A_1$	$C_1$
$A_2$	$C_2$
$A_3$	$C_3$
$\vdots$	$\vdots$
$A_N$	$C_N$

B

モデル	$C_1$	$C_2$	$C_3$	-	-	-	-	-	-	$C_N$
$M_1$	$P_5$	$P_{10}$	$P_1$	-	-	-	-	-	-	$P_4$
$M_2$	$P_{21}$	$P_3$	$P_8$	-	-	-	-	-	-	$P_{24}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$							$\vdots$
$M_M$	$P_{13}$	$P_2$	$P_{19}$	-	-	-	-	-	-	$P_7$

C

状態 モデル	①	②	③	④
$M_1$	$V_3, V_5, V_6$	$V_1, V_8, V_{12}$	$V_8, V_{11}$	$V_2, V_{10}, V_{13}$
$M_2$	$V_1, V_5$	$V_4, V_7, V_8$	$V_6, V_9, V_{15}$	$V_3, V_{17}$
$\vdots$				
$M_M$	$V_9, V_{10}, V_{13}$	$V_1, V_5$	$V_8, V_9, V_{14}$	$V_1, V_2, V_7$

【図 6】

図 6

状態 a

モデル \ 分布	V <sub>1</sub>	V <sub>2</sub>	V <sub>3</sub>	- - -	V <sub>N</sub>
M <sub>1</sub>	W <sub>11</sub>	W <sub>8</sub>	W <sub>1</sub>	- - -	W <sub>5</sub>
M <sub>2</sub>	W <sub>2</sub>	W <sub>4</sub>	W <sub>15</sub>	- - -	W <sub>7</sub>
⋮	⋮	⋮	⋮		⋮
M <sub>M</sub>	W <sub>16</sub>	W <sub>3</sub>	W <sub>7</sub>	- - -	W <sub>21</sub>

状態 b

モデル \ 分布	V <sub>1</sub>	V <sub>2</sub>	- - -	V <sub>N</sub>
M <sub>1</sub>	-	-	-	-
M <sub>2</sub>	-	-	-	-
⋮				
M <sub>M</sub>	-	-	-	-

状態 c

モデル \ 分布	-	-	-	-	-
	-	-	-	-	-
	-	-	-	-	-

状態 d

モデル \ 分布	-	-	-	-	-
	-	-	-	-	-
	-	-	-	-	-